Priority Research Direction:
Secure Federation for Large-Scale Collaborative Science

Key emerging challenges

***ASCR security technologies underpin science federation projects worldwide. But now we face*:**

- **More people: Collaborations larger, more frequent, more fluid, span more institutions**
- **More resources: Large and diverse data, computers, sensors, mobile devices**
- **More attacks: More, better connected, smarter, better equipped adversaries**
- **New technologies: Evolving commodity tech**

Summary of research direction

***Rethink security enforcement, intrusion detection, incident response for team science*:**

- **Enable instant VOs, that provision secure collaborative spaces in minutes not weeks**
- **Leverage VOs as basis for security mechanisms**
- **Take systems perspective on DOE-wide security, integrating data and modeling**
- **Address challenges of massively instrumented and mobile environment**

Potential impact on software/systems

- **"Virtual private/team labs" via outsourced IT provide delivery and enforcement vehicle; require delivery and sustainability model**
- **Data architectures and their security must be rethought to enable secure but flexible sharing**
- **HPC analysis and modeling in support of computer security → link to applied math**

Potential impact on science communities or DOE capabilities

- **Intrinsic and proactive security mechanisms encourage rather than discourage collaboration**
- **Virtual private labs enable secure creation and operation of outsourced research IT**
- **By facilitating secure data sharing spanning many modalities, enable transparent science**
- **Deliver value to 26,000 DOE/SC researchers and 27,000 facility users**

Priority Research Direction: High-Performance Distributed Data Management

Key emerging challenges

ASCR technologies have made reliable, efficient, secure data movement routine. But now we face:

- **Big pipes: 100 Gbps will soon be commonplace, and 1 Tbps networks are on the horizon**
- **Massive data: Data volumes and complexities are growing faster than networks, computers**
- **More demands: Data becomes central to research progress; more sophisticated data management and sharing capabilities needed**

Summary of research direction

A coordinated, multi-dimensional attack on big data and its place in the research ecosystem:

1) **End-to-end, Tbps file system-to-file system, with management of large-scale parallelism**
2) **Top-to-bottom, from problem to answer, addressing need for services beyond movement**
3) **System-wide, with 1000s of 10+ Gbps users, encompassing large-scale automation**

Potential impact on software/systems

- **Data services recognized as facility-supported service: ubiquitous, secure, robust**
- **Data services expand to encompass not just movement but also storage, analysis, indexing, sharing, annotation, …**
- **Co-design of file systems, storage systems, operating systems, data movement systems, networks, applications**

Potential impact on science communities or DOE capabilities

- **Data pipes never an unnecessary barrier to use of DOE high-end facilities**
- **Data reuse and pace of discovery both increase by order of magnitude or more**
- **DOE facilities are used more efficiently due to more effective engagement of users with tasks**
- **Number of facility users increases significantly**

Priority Research Direction:
Virtual Private Labs for Extreme Science

Key emerging challenges

- Increasingly complex science projects involving large-scale collaboration, data, and workflows require many tools from many sources
- Individual projects can not feasibly assemble these tools into a coherent and usable whole
- Challenges exacerbated by security, data issues
- Sustainability for the increasingly complex and expensive software required for science success

Summary of research direction

- Enable the creation of virtual private labs that provide the resources and tools needed for effective extreme-scale team science
- Define platform that facilitates application development, distribution, operation, and use
- Address challenges inherent in massive data and security
- Do for extreme science what Apple IOS and app store does for the consumer market

Potential impact on software/systems

- New type of facility required to support outsourced collaborative science applications
- New approach to developing and delivering collaborative applications lowers costs of producing, delivering, and using those apps
- New security challenges relating to massive outsourcing and collaboration
- Need for co-design of data architecture and collaboration architecture

Potential impact on science communities or DOE capabilities

- Acceleration of on-ramp/off-ramp activities around extreme-scale computing reduces time to discovery and increases access to results
- Dramatic improvement in quality of research tools and access to computing resources accelerates discovery across DOE in general